Consensus optimization on manifolds

Alain Sarlette   Rodolphe Sepulchre*

November 26, 2008

Abstract

The present paper considers distributed consensus algorithms that involve N agents evolving on a connected compact homogeneous manifold. The agents track no external reference and communicate their relative state according to a communication graph. The consensus problem is formulated in terms of the extrema of a cost function. This leads to efficient gradient algorithms to synchronize (i.e. maximizing the consensus) or balance (i.e. minimizing the consensus) the agents; a convenient adaptation of the gradient algorithms is used when the communication graph is directed and time-varying. The cost function is linked to a specific centroid definition on manifolds, introduced here as the *induced arithmetic mean*, that is easily computable in closed form and may be of independent interest for a number of manifolds. The special orthogonal group $SO(n)$ and the Grassmann manifold $Grass(p, n)$ are treated as original examples. A link is also drawn with the many existing results on the circle.

1 Introduction

The distributed computation of means/averages of datasets (in an algorithmic setting) and the *synchronization* of a set of agents (in a control setting) — i.e. driving all the agents to a common point in state space — are ubiquitous tasks in current engineering problems. Likewise, spreading a set of agents in the available state space — linked to the definition of *balancing* in §4 — is a classical problem of growing interest. Practical applications include autonomous swarm/formation operation (e.g. [42; 28; 22; 23; 25]), distributed decision making (e.g. [35; 47]), neural and communication networks (e.g. [46; 19]), clustering and other reduction methods (e.g. [17]), optimal covering or coding (e.g. [3; 4; 11; 12]) and other fields where averaging/synchronizing or distributing a set of points appear as sub-problems. In a modeling framework, the understanding of synchronization or more generally swarm behavior has also led to many important studies (e.g. [26; 45; 48]).

Synchronization algorithms are well understood in Euclidean spaces (e.g. [33; 32; 47; 35]). They are based on the natural definition and distributed computation of the centroid in $\mathbb{R}^m$. However, many applications above involve manifolds that are not homeomorphic to an Euclidean space. Even for formations moving in $\mathbb{R}^2$ or $\mathbb{R}^3$, the agents' orientations evolve in a manifold $SO(2) \cong S^1$ or $SO(3)$. Balancing only makes sense on compact state spaces; though many theoretical results concern convex or star-shaped subsets of $\mathbb{R}^m$ (e.g. [12]), most applications

---

*Department of Electrical Engineering and Computer Science (Montefiore Institute), University of Liège, Sart-Tilman Bldg. B28, B-4000 Liège, Belgium (`alain.sarlette@ulg.ac.be, r.sepulchre@ulg.ac.be`).

involve compact manifolds. It seems that the study of global synchronization or balancing in non-Euclidean manifolds is not widely covered in the literature, except for the circle.

The present paper proposes algorithms for global synchronization and balancing — grouped under the term *consensus* — on connected compact homogeneous manifolds. A homogeneous manifold $\mathcal{M}$ is isomorphic to the quotient of two Lie groups. Intuitively, it is a manifold on which "all points are equivalent". This makes the problem symmetric with respect to the absolute position on the manifold and allows to focus on *configurations* of the swarm, i.e. *relative positions* of the agents.

The main idea is to embed $\mathcal{M}$ in $\mathbb{R}^m$ and measure distances between agents in $\mathbb{R}^m$ in order to build a convenient cost function for an optimization-based approach. The related centroid on $\mathcal{M}$ may be interesting on its own account; it is therefore studied in more detail in §3.

Throughout the paper, the abstract concepts are illustrated on the special orthogonal group $SO(n)$, the Grassmann manifold $Grass(p,n)$ of p-dimensional vector spaces in $\mathbb{R}^n$ and sometimes the circle S^1, which is in fact isomorphic to both $SO(2)$ and $Grass(1,2)$. Other manifolds to which the present framework could be applied include the n-dimensional spheres S^n and the connected compact Lie groups. The circle S^1 is the simplest example; it links the present work to existing results in [42; 41; 39; 43]. $SO(n)$ is important in control applications as the natural state space for orientations of n-dimensional rigid bodies. $Grass(p,n)$ rather appears in algorithmic problems; [11] mentions the optimal placement of N laser beams for cancer treatment and the projection of multi-dimensional data on N representative planes as practical applications of optimal distributions on $Grass(p,n)$.

The paper is organized as follows. Previous work is briefly reviewed in §1.1. Section 2 introduces concepts and notations about graph theory, $SO(n)$ and $Grass(p,n)$. Section 3 is devoted to the induced arithmetic mean. A definition of consensus is presented in §4. Section 5 introduces a cost function to express the consensus problem in an optimization setting. Section 6 derives gradient algorithms based on this cost function, the only communicated information being the relative positions of interconnected agents; convergence is proved for any connected, fixed and undirected communication graph. Algorithms whose convergence properties can be guaranteed under possibly directed, time-varying and disconnected communication graphs are presented in §7; they employ an auxiliary variable that evolves in the embedding space $\mathbb{R}^m$.

1.1   Previous work

Most of the work related to synchronization and balancing on manifolds concerns the circle S^1. The most extensive literature on the subject derives from the Kuramoto model (see [44] for a review). Recently however, synchronization on the circle has been considered from a control perspective, the state variables representing headings of agents in the plane. Most results cover *local* convergence [22; 33]. An interesting set of *globally* convergent algorithms in $SE(2) = S^1 \times \mathbb{R}^2$ is presented in [42], but they require all-to-all communication. Some problems related to global discrete-time synchronization on S^1 under different communication constraints are discussed in [37], where connections of the control problem with various existing models are made. Stronger results are presented in [39] for global synchronization and balancing on S^1 with varying, directed communication links, at the cost of introducing estimator variables that communicating agents must exchange. Finally, [43] presents results on $SE(2)$ similar to those of [42] but under relaxed communication assumptions, using among others the estimator strategy of [39; 40].

Several authors have already presented algorithms that asymptotically synchronize satellite attitudes, involving the rotation group $SO(3)$. They often rely on tracking a common external reference (e.g. [27]) or leader (e.g. [5; 25]). The use of the convenient but non-unique quaternion representation for $SO(3)$ produces unwanted artefacts in the satellites' motions. Attitude synchronization without common references and quaternion artefacts is studied in [34]; using the same distance measure as the present work, an artificial coupling potential is built to establish local stability. All these approaches explicitly incorporate the second-order rigid-body dynamics. In accordance with the consensus approach, the present paper reduces the agents to first-order kinematic models to focus on (almost) global convergence for various agent interconnections, without

any leader or external reference. Application of this framework in a mechanical setting is discussed in a separate paper [38].

Synchronization or balancing on a manifold $\mathcal{M}$ is closely related to the definition and computation of a mean or centroid of points on $\mathcal{M}$, a basic problem that has attracted somewhat more attention, as can be seen from [20; 9; 15] among others.

A key element of the present paper is the computation of a centroid in the embedding space $\mathbb{R}^m$ of $\mathcal{M}$, which is then projected onto $\mathcal{M}$. This is connected to the "projected arithmetic mean" defined in [31] for $SO(3)$. In fact, the computation of statistics in a larger and simpler embedding manifold (usually Euclidean space) and projecting the result back onto the original manifold, goes back to 1972 [13].

A short example in [1] addresses the computation of a "centroid of subspaces", without much theoretical analysis; in fact, our algorithms on $Grass(p, n)$ are similar and can eventually be viewed as generalizing the developments in [1] in the framework of consensus and synchronization. More recently, [17] uses the centroid associated to the projector representation of $Grass(p, n)$, exactly as is done below but without going into theoretical details, to compute the cluster centers in a clustering algorithm. The distance measure associated to this centroid on $Grass(p, n)$ is called the *chordal distance* in [11; 4] where it is used to derive optimal distributions ("packings") of N agents on some specific Grassmann manifolds.

Finally, the topic of optimization-based algorithm design on manifolds has considerably developed over the last decades (see e.g. [6], [14] and the books [18; 2]).

2   Preliminaries

2.1   Elements of graph theory

Consensus among a group of agents depends on the available communication links. When considering limited agent interconnections, it is customary to represent communication links by means of a *graph*. The graph G is composed of N *vertices* (the N agents) and contains the *edge* (j, k) if agent j sends information to agent k, which is denoted $j \rightsquigarrow k$. A positive *weight* a_{jk} is associated to each edge (j, k) to obtain a weighted graph; the weight is extended to any pair of vertices by imposing $a_{jk} = 0$ iff (j, k) does not belong to the edges of G. The full notation for the resulting *digraph* (directed graph) is $G(V, E, A)$ where $V = \{\text{vertices}\}$, $E = \{\text{edges}\}$ and A, containing the a_{jk}, is the *adjacency matrix*. The convention $a_{kk} = 0$ $\forall k$ is assumed for the representation of communication links.

The *out-degree* of a vertex k is the quantity $d_k^{(o)} = \sum_{j=1}^{N} a_{kj}$ of information leaving k towards other agents; its *in-degree* is the quantity $d_k^{(i)} = \sum_{j=1}^{N} a_{jk}$ of information received by k from other agents. These degrees can be assembled in diagonal matrices $D^{(o)}$ and $D^{(i)}$. A graph is *balanced* if $D^{(o)} = D^{(i)}$. This is satisfied in particular by *undirected* graphs, for which $A = A^T$. A graph is *bidirectional* if $(j, k) \in E \Leftrightarrow (k, j) \in E$ (but not necessarily $A = A^T$).

The *Laplacian L* of a graph is $L = D - A$. For directed graphs, $D^{(i)}$ or $D^{(o)}$ can be used, leading to the in-Laplacian $L^{(i)} = D^{(i)} - A$ and the out-Laplacian $L^{(o)} = D^{(o)} - A$. By construction, $L^{(i)}$ has zero column sums and $L^{(o)}$ has zero row sums. The spectrum of the Laplacian reflects several interesting properties of the associated graph, specially in the case of undirected graphs (see for example [10]).

$G(V, E, A)$ is *strongly connected* if it contains a directed path from any vertex j to any vertex l (i.e. a sequence of vertices starting with j and ending with l such that $(v_k, v_{k+1}) \in E$ for any two consecutive vertices v_k, v_{k+1}); G is *weakly connected* if there is such a path in the *associated undirected graph*, with adjacency matrix $A + A^T$.

For time-varying interconnections, a time-varying graph $G(t)$ is used and all the previously defined elements simply depend on time. If the elements of $A(t)$ are bounded and satisfy some threshold $a_{jk}(t) \geq \delta > 0$ $\forall (j, k) \in E(t)$ and $\forall t$, then $G(t)$ is called a *δ-digraph*. The present paper

always considers δ-digraphs.

In a δ-digraph $G(V, E, A)$, vertex j is said to be *connected* to vertex k *across* $[t_1, t_2]$ if there is a path from j to k in the digraph $\bar{G}(V, \bar{E}, \bar{A})$ defined by

$$\bar{a}_{jk} = \begin{cases} \int_{t_1}^{t_2} a_{jk}(t)dt & \text{if } \int_{t_1}^{t_2} a_{jk}(t)dt \geq \delta \\ 0 & \text{if } \int_{t_1}^{t_2} a_{jk}(t)dt < \delta \end{cases}$$

$$(j, k) \in \bar{E} \text{ iff } \bar{a}_{jk} \neq 0 .$$

$\bar{G}$ can be seen as a time-integrated graph while the δ-criterion prevents vanishing edges. A δ-digraph $G(t)$ is called *uniformly connected* if there exist a vertex k and a time horizon $T > 0$ such that $\forall t$, k is connected to all other vertices across $[t, t + T]$.

2.2   Specific manifolds

The concepts presented in this paper are illustrated on two particular manifolds: $SO(n)$ and $Grass(p, n)$.

The special orthogonal Lie group $SO(n)$ It can be viewed as the set of positively oriented orthonormal bases of $\mathbb{R}^n$, or equivalently as the set of rotation matrices in $\mathbb{R}^n$; it is the natural state space for the orientation of a rigid body in $\mathbb{R}^n$. In its canonical representation, used in the present paper, a point of $SO(n)$ is characterized by a real $n \times n$ orthogonal matrix Q with determinant equal to $+1$. $SO(n)$ is homogeneous (as any Lie group), compact and connected. It has dimension $n(n-1)/2$.

The Grassmann manifold $Grass(p, n)$ Each point on $Grass(p, n)$ denotes a p-dimensional subspace $\mathcal{Y}$ of $\mathbb{R}^n$. The dimension of $Grass(p, n)$ is $p(n-p)$. Since $Grass(n-p, n)$ is isomorphic to $Grass(p, n)$ by identifying orthogonally complementary subspaces, this paper assumes without loss of generality that $p \leq \frac{n}{2}$. For the special case $p = 1$, the Grassmann manifold $Grass(1, n)$ is also known as the *projective space* in dimension n. $Grass(p, n)$ is connected, compact and homogeneous as the quotient of the orthogonal Lie group $O(n)$ by $O(p) \times O(n-p)$. Indeed, $\mathcal{Y} \in Grass(p, n)$ can be represented for instance by a (not necessarily positively oriented) orthonormal basis $Q \in O(n)$ whose first p column-vectors span $\mathcal{Y}$; the same point $\mathcal{Y} \in Grass(p, n)$ is represented by any Q whose first p column-vectors span $\mathcal{Y}$ ($O(p)$-symmetry) and whose last $n - p$ column-vectors span the orthogonal complement of $\mathcal{Y}$ ($O(n-p)$-symmetry). Other quotient structures for $Grass(p, n)$ are discussed in [1].

A matrix manifold representation of $Grass(p, n)$ found in [1] assigns to $\mathcal{Y}$ any $n \times p$ matrix Y of p orthonormal column-vectors spanning $\mathcal{Y}$ (p-basis representation); all Y corresponding to rotations and reflections of the p column-vectors in $\mathcal{Y}$ represent the same $\mathcal{Y}$ ($O(p)$-symmetry), so this representation is not unique. The dimension of this representation is $np - p(p + 1)/2$. In [29], a point of $Grass(p, n)$ is represented by $\Pi = YY^T$, the orthonormal projector on $\mathcal{Y}$ (projector representation); using the orthonormal projector on the space orthogonal to $\mathcal{Y}$, $\Pi_\perp = I_n - YY^T$ where I_n denotes the $n \times n$ identity matrix, is strictly equivalent. The main advantage of this representation is that there exists a bijection between $Grass(p, n)$ and the orthonormal projectors of rank p, such that the projector representation makes $Grass(p, n)$ an embedded submanifold of the cone $\mathbb{S}_n^+$ of $n \times n$ symmetric positive semi-definite matrices. A disadvantage of this representation is its large dimension $n(n + 1)/2$.

3   The induced arithmetic mean

A homogeneous manifold $\mathcal{M}$ is a manifold with a transitive group action by a Lie group $\mathcal{G}$: it is isomorphic to the quotient manifold $\mathcal{G}/\mathcal{H}$ of a group $\mathcal{G}$ by one of its subgroups $\mathcal{H}$. Informally,

it can be seen as a manifold on which "all points are equivalent". The present paper considers connected compact homogeneous manifolds satisfying the following embedding property.

Assumption 1 *$\mathcal{M}$ is a connected compact homogeneous manifold smoothly embedded in $\mathbb{R}^m$ with the Euclidean norm $\|y\| = r_{\mathcal{M}}$ constant over $y \in \mathcal{M}$. The Lie group $\mathcal{G}$ acts as a subgroup of the orthogonal group on $\mathbb{R}^m$.*

It is a well-known fact of differential geometry that any smooth $\frac{m}{2}$-dimensional Riemannian manifold can be smoothly embedded in $\mathbb{R}^m$. The additional condition $\|y\| = r_{\mathcal{M}}$ is in agreement with the fact that all points on $\mathcal{M}$ should be equivalent. It is sometimes preferred to represent $y \in \mathcal{M}$ by a matrix $B \in \mathbb{R}^{n_1 \times n_2}$ instead of a vector. Componentwise identification $\mathbb{R}^{n_1 \times n_2} \cong \mathbb{R}^m$ is assumed whenever necessary; the corresponding norm is the *Frobenius norm* $\|B\| = \sqrt{\text{trace}(B^T B)}$.

Consider a set of N agents on a manifold $\mathcal{M}$ satisfying Assumption 1. The position of agent k is denoted by y_k and its weight by w_k.

Definition 1 *The **induced arithmetic mean** IAM $\subseteq \mathcal{M}$ of N agents of weights $w_k > 0$ and positions $y_k \in \mathcal{M}$, $k = 1...N$, is the set of points in $\mathcal{M}$ that globally minimize the weighted sum of squared Euclidean distances in $\mathbb{R}^m$ to each y_k:*

$$IAM \;=\; \operatorname*{argmin}_{c \in \mathcal{M}} \textstyle\sum_{k=1}^{N} w_k \, d_{\mathbb{R}^m}^2(y_k, c) \;=\; \operatorname*{argmin}_{c \in \mathcal{M}} \textstyle\sum_{k=1}^{N} w_k \, (y_k - c)^T (y_k - c) \;. \tag{1}$$

*The **anti-[induced arithmetic mean]** AIAM $\subseteq \mathcal{M}$ is the set of points in $\mathcal{M}$ that globally maximize the weighted sum of squared Euclidean distances in $\mathbb{R}^m$ to each y_k:*

$$AIAM \;=\; \operatorname*{argmax}_{c \in \mathcal{M}} \textstyle\sum_{k=1}^{N} w_k \, d_{\mathbb{R}^m}^2(y_k, c) \;=\; \operatorname*{argmax}_{c \in \mathcal{M}} \textstyle\sum_{k=1}^{N} w_k \, (y_k - c)^T (y_k - c) \;. \tag{2}$$

The terminology is derived from [31] where the *IAM* on $SO(3)$ is called the *projected arithmetic mean*. The point in Definition 1 is that distances are measured *in the embedding space* $\mathbb{R}^m$. It thereby differs from the canonical definition of mean of N agents on $\mathcal{M}$, the *Karcher mean* [24; 36; 16; 20], which uses the geodesic distance $d_{\mathcal{M}}$ along the Riemannian manifold $\mathcal{M}$ (with, in the present setting, the Riemannian metric induced by the embedding of $\mathcal{M}$ in $\mathbb{R}^m$):

$$C_{Karcher} = \operatorname*{argmin}_{c \in \mathcal{M}} \; \textstyle\sum_{k=1}^{N} w_k \, d_{\mathcal{M}}^2(y_k, c) \;.$$

The induced arithmetic mean has the following properties.

1. The *IAM* of a single point y_1 is the point itself.

2. The *IAM* is invariant under permutations of agents of equal weights.

3. The *IAM* commutes with the symmetry group of the homogeneous manifold.

4. The *IAM* does not always reduce to a single point.

The last feature seems unavoidable for any mean (including the Karcher mean) that satisfies the other properties. The main advantage of the *IAM* over the Karcher mean is computational. The *IAM* and *AIAM* are closely related to the centroid.

Definition 2 *The **centroid** $C_e \in \mathbb{R}^m$ of N weighted agents located on $\mathcal{M}$ is*

$$C_e = \tfrac{1}{W} \textstyle\sum_{k=1}^{N} w_k \, y_k \;, \qquad where \;\; W = \textstyle\sum_{k=1}^{N} w_k \;.$$

Since $\|c\| = r_{\mathcal{M}}$ for $c \in \mathcal{M}$ by Assumption 1, equivalent definitions for the *IAM* and *AIAM* are

$$IAM = \underset{c \in \mathcal{M}}{\mathrm{argmax}}(c^T C_e) \qquad \text{and} \qquad AIAM = \underset{c \in \mathcal{M}}{\mathrm{argmax}}(-c^T C_e) \, . \tag{3}$$

Hence, computing the *IAM* and *AIAM* just involves a search for the global maximizers of a linear function on $\mathbb{R}^m$ in a very regular search space $\mathcal{M}$. Local maximization methods would even suffice if the linear function had no maxima on $\mathcal{M}$ other than the global maxima. This is the case for any linear function on $SO(n)$ and $Grass(p, n)$ (see §3.1) as well as the n-dimensional sphere S^n in $\mathbb{R}^{n+1}$. Not knowing whether this property holds for all manifolds satisfying Assumption 1, we formulate the following blanket assumption.

Assumption 2 *The local maxima of any linear function $f(c) = c^T b$ over $c \in \mathcal{M}$, with b fixed in $\mathbb{R}^m$, are all global maxima.*

3.1   Examples

These examples exclusively consider the *IAM*; from (3), the conclusions for the *AIAM* are simply obtained by replacing C_e with $-C_e$.

The circle   The circle embedded in $\mathbb{R}^2$ with its center at the origin satisfies Assumptions 1 and 2. The *IAM* is simply the central projection of C_e onto the circle. Hence it corresponds to the whole circle if $C_e = 0$ and reduces to a single point in other situations. The *IAM* uses the chordal distance between points, while the Karcher mean would use arclength distance.

The special orthogonal group   The embedding of $SO(n)$ as orthogonal matrices $Q \in \mathbb{R}^{n \times n}$, $\det(Q) > 0$, satisfies Assumption 1 since $\|Q\| = \sqrt{\mathrm{trace}(Q^T Q)} = \sqrt{n}$. It also satisfies Assumption 2 (proof in §6). $C_e = \sum_k Q_k$ is a general $n \times n$ matrix. The *IAM* is linked to the *polar decomposition* of C_e. Any matrix B can be decomposed into UR with U orthogonal and R symmetric positive semi-definite; R is always unique, U is unique if B is non-singular [7]. Each U is a global minimizer of $d_{\mathbb{R}^{n \times n}}(c, B)$ over $c \in O(n)$. Thus, if $\det(C_e) \geq 0$, the *IAM* contains all matrices $U \colon \det(U) > 0$ obtained from the polar decomposition of C_e; this was already noticed in [31]. When $\det(C_e) < 0$, the result is more complicated but still has a closed-form solution.

Proposition 1 *Consider U an orthogonal matrix obtained from the polar decomposition $C_e = UR$. The IAM of N points on $SO(n)$ is characterized as follows.*

> *If $\det(C_e) \geq 0$, then $IAM = \{U : \det(U) > 0\}$. It reduces to a single point if the multiplicity of 0 as an eigenvalue of C_e is less or equal to 1.*

> *If $\det(C_e) \leq 0$, then $IAM = \{U H J H^T\}$ where $\det(U) < 0$, H contains the orthonormalized eigenvectors of R with an eigenvector corresponding to the smallest eigenvalue of R in the first column, and $J = \begin{pmatrix} -1 & 0 \\ 0 & I_{n-1} \end{pmatrix}$. The IAM reduces to a single point if the smallest eigenvalue of R has multiplicity 1.*

Proof: It is provided in §6 after introducing further necessary material to compute the critical points of $c^T C_e$, among which the local maxima are selected.   □

The Grassmann manifold The representation of $Grass(p, n)$ with p-bases Y_k is not an embedding and cannot be used in the proposed framework, because the p-dimensional subspace of $\mathbb{R}^n$ spanned by the columns of $C_e = \sum_k Y_k$ would depend on the particular matrices Y_k chosen to represent the subspaces $\mathcal{Y}_k$. The *IAM* is defined with the projector representation, embedding $Grass(p, n)$ in $\mathbb{S}_n^+$. The latter satisfies Assumption 1; the Frobenius norm of a p-rank projector is $\sqrt{p}$. It also satisfies Assumption 2 (proof in §6). The centroid C_e of N projectors is generally a symmetric positive semi-definite matrix of rank $\geq p$.

Proposition 2 *The IAM contains all dominant p-eigenspaces of C_e. It reduces to a single point if the p-largest and $(p + 1)$-largest eigenvalues of C_e are different.*

Proof: Following the same lines as for $SO(n)$, it is postponed to §6.   □

In fact, for $\mathcal{Y} \in Grass(p, n)$ with a p-basis Y and the projector $\Pi_{\mathcal{Y}} = YY^T$, the cost function in (3) becomes

$$f(\Pi_{\mathcal{Y}}) = \text{trace}(\Pi_{\mathcal{Y}} C_e) = \text{trace}(Y^T C_e Y) = \text{trace}((Y^T Y)^{-1} Y^T C_e Y) \tag{4}$$

where the last expression is equal to the generalized Rayleigh quotient for the computation of the dominant p-eigenspace of C_e. The computation of eigenspaces from cost function (4) is extensively covered in [1; 2]. Furthermore, it is a well-known fact of linear algebra that the p largest eigenvalues (the others being 0) of $\Pi_{\mathcal{Y}} \Pi_k$ are the squared cosines of the principal angles ϕ_k^i, $i = 1...p$, between subspaces $\mathcal{Y}$ and $\mathcal{Y}_k$. This provides a geometrical meaning for the *IAM* of subspaces: it minimizes the sum of squared sines of principal angles between the set of subspaces $\mathcal{Y}_k$, $k = 1...N$, and a centroid candidate subspace $\mathcal{Y}$, i.e. $IAM = \text{argmin}_{\mathcal{Y}} \sum_{k=1}^N \sum_{i=1}^p \sin^2(\phi_k^i)$. The Karcher mean admits the same formula with $\sin^2(\phi_k^i)$ replaced by $(\phi_k^i)^2$ [11].

4   Consensus

Consider a set of agents with positions y_k, $k = 1...N$, on a manifold $\mathcal{M}$ satisfying Assumption 1. The rest of this paper assumes equal weights $w_k = 1$ $\forall k$; extension to weighted agents is straightforward. Suppose that the agents are interconnected according to a fixed digraph G of adjacency matrix $A = [a_{jk}]$.

Definition 3 *Synchronization is the configuration where $y_j = y_k$ $\forall j, k$.*

Definition 4 *A **consensus** configuration with graph G is a configuration where each agent k is located at a point of the IAM of its neighbors $j \rightsquigarrow k$, weighted according to the strength of the corresponding edge. Similarly, an **anti-consensus** configuration satisfies this definition with IAM replaced by AIAM.*

$$\textbf{\textit{Consensus:}} \qquad y_k \in \underset{c \in \mathcal{M}}{\text{argmax}} \left(c^T \sum_{j=1}^N a_{jk} \, y_j \right) \qquad \forall k \; . \tag{5}$$

$$\textbf{\textit{Anti-consensus:}} \qquad y_k \in \underset{c \in \mathcal{M}}{\text{argmin}} \left(c^T \sum_{j=1}^N a_{jk} \, y_j \right) \qquad \forall k \; . \tag{6}$$

Note that consensus is defined as a Nash equilibrium: each agent minimizes its cost function *assuming the others fixed*; the possibility to decrease cost functions by moving several agents simultaneously is not considered. Consensus is graph-dependent: agent k reaches consensus when it minimizes its distance to agents $j \rightsquigarrow k$.

Proposition 3 *If G is an equally-weighted complete graph, then the only possible consensus configuration is synchronization.*

Proof:At consensus the y_k satisfy $\quad y_k^T \sum_{j \neq k} y_j \geq c^T \sum_{j \neq k} y_j \quad \forall k$ and $\forall c \in \mathcal{M}$. Furthermore, it is obvious that $y_k^T y_k > y_k^T c$ for any $c \in \mathcal{M} \setminus \{y_k\}$. As a consequence, $y_k^T \sum_{j=1}^N y_j > c^T \sum_{j=1}^N y_j \quad \forall c \in \mathcal{M} \setminus \{y_k\}$ and $\forall k$. Thus according to (3), each y_k is located at the *IAM* of all the agents, which moreover reduces to a single point; thus $y_k = y_j = IAM(\{y_l : l = 1...N\}) \, \forall k, j$.
$\square$

Synchronization is a configuration of complete consensus. To similarly characterize a configuration of complete anti-consensus, it appears meaningful to require that the *IAM* of the agents is the entire manifold $\mathcal{M}$; this is called a *balanced* configuration.

Definition 5 *N agents are **balanced** if their IAM contains all $\mathcal{M}$.*

Balancing implies some spreading of the agents on the manifold. A full characterization of balanced configurations seems complicated. Balanced configurations do not always exist (typically, when the number of agents is too small) and are mostly not unique (they can appear in qualitatively different forms). The following link exists between anti-consensus for the equally-weighted complete graph and balancing.

Proposition 4 *All balanced configurations are anti-consensus configurations for the equally-weighted complete graph.*

Proof: For the equally-weighted complete graph, (6) can be written

$$y_k \in \operatorname*{argmin}_{c \in \mathcal{M}} \left(c^T \left(N \, C_e - y_k \right) \right) \qquad \forall k \; . \tag{7}$$

Assume that the agents are balanced. This means that $f(c) = c^T C_e$ must be constant over $c \in \mathcal{M}$. Therefore (7) reduces to $y_k = y_k \, \forall k$ which is trivially satisfied. $\square$

In contrast to Proposition 3, Proposition 4 does not establish a necessary and sufficient condition; and indeed, anti-consensus configurations for the equally-weighted complete graph that are not balanced do exist, though they seem exceptional.

4.1   Examples

The following examples illustrate among others the last assertions about balanced configurations.

The circle Anti-consensus configurations for the equally-weighted complete graph are fully characterized in [42]. It is shown that the only anti-consensus configurations that are not balanced correspond to $(N + 1)/2$ agents at one position and $(N - 1)/2$ agents at the opposite position on the circle, for N odd. Balanced configurations are unique for $N = 2$ and $N = 3$ and form a continuum for $N > 3$.

Another interesting illustration is the equally-weighted *undirected ring graph* in which each agent is connected to two neighbors such that the graph forms a single closed undirected path. Regular consensus configurations correspond to situations with consecutive agents in the path always separated by the same angle $0 \leq \chi \leq \pi/2$; regular anti-consensus configurations have $\pi/2 \leq \chi \leq \pi$. In addition, for $N \geq 4$, irregular consensus and anti-consensus configurations exist where non-consecutive angles of the regular configurations are replaced by $(\pi - \chi)$. As a consequence:

1. Several qualitatively different (anti-)consensus configurations exist.

2. Consensus and anti-consensus configurations can be equivalent when discarding the graph. For example, the positions occupied by 7 agents separated by $2\pi/7$ (consensus) or $4\pi/7$ (anti-consensus) are strictly equivalent; the only difference, based on *which agent* is located at *which position*, concerns the way the links are drawn.

3. Degenerate configurations of simultaneous consensus and anti-consensus exist (e.g. $\chi = \pi/2$ for $N = 4, 8, ...$); this singularity is specific to the particular graph.

4. There is no common anti-consensus state for all ring graphs. Indeed, considering an agent k, a common anti-consensus state would require that any two other agents, as potential neighbors of k, are either separated by π or located at both sides of k at a distance $\chi \geq \pi/2$; one easily verifies that this cannot be satisfied for all k.

The special orthogonal group   Simulations of the algorithms proposed in this paper suggest that balanced configurations always exist for $N \geq 2$ if n is even and for $N \geq 4$ if n is odd. Under these conditions, convergence to an anti-consensus state that is not balanced is not observed for the equally-weighted complete graph.

The Grassmann manifold   Balanced states on $Grass(p, n)$ appear if all eigenvalues of C_e are equal. Since $\mathrm{trace}(C_e) = \frac{1}{N} \sum_k \mathrm{trace}(\Pi_k) = p$, this requires $C_e = \frac{p}{n} I_n$. This is not always possible with N orthonormal projectors of rank p. As for $SO(n)$, simulations tend to indicate that it is possible when N is large enough; however, computing the minimal value of N for a given n and p is not straightforward.

5   Consensus optimization strategy

The presence of a maximization condition in the definitions of the previous sections naturally points to the use of optimization methods. The present section introduces a cost function whose optimization leads to (anti-)consensus configurations. For a graph G with adjacency matrix $A = [a_{jk}]$ and associated Laplacian $L^{(i)} = [l_{jk}^{(i)}]$ and the variable $y = (y_1, ..., y_N) \in \mathcal{M}^N$, define

$$P_L(y) \;=\; \tfrac{1}{2N^2}\sum_{k=1}^{N}\sum_{j=1}^{N} a_{jk}\, y_j^T y_k \;=\; \xi_1 - \tfrac{1}{4N^2}\sum_{k=1}^{N}\sum_{j=1}^{N} a_{jk}\, \|y_j - y_k\|^2 \qquad (8)$$

with constant $\xi_1 = \frac{r_{\mathcal{M}}^2}{4N^2}\sum_k \sum_j a_{jk}$. The index L refers to the fact that (8) can also be written as a quadratic form on the graph Laplacian:

$$P_L(y) \;=\; \xi_2 - \tfrac{1}{2N^2}\sum_{k=1}^{N}\sum_{j=1}^{N} l_{jk}^{(i)}\, y_j^T y_k \qquad \text{with constant } \xi_2 = \frac{r_{\mathcal{M}}^2}{2N^2}\sum_k d_k^{(i)}. \qquad (9)$$

In [37] and [43], this form of P_L is studied on the circle for undirected equally-weighted graphs. For the unit-weighted complete graph, $P := P_L + \frac{r_{\mathcal{M}}^2}{2N}$ equals

$$P(y) = \tfrac{1}{2}\|C_e\|^2 , \qquad (10)$$

proportional to the squared norm of the centroid C_e. This is a classical measure of the synchrony of phase variables on the circle S^1, used for decades in the literature on coupled oscillators; in the context of the Kuramoto model, $P(y)$ is known as the "complex order parameter" (because $\mathbb{R}^2$ is usually identified with $\mathbb{C}$ in that context). In [42], P is used to derive gradient algorithms for synchronization (by maximizing (10)) or balancing (by minimizing (10)) on S^1.

Proposition 5 *Synchronization of the N agents on $\mathcal{M}$ is the unique global maximum of P_L whenever the graph G associated to $L^{(i)}$ is weakly connected.*

Proof: According to the second form of (8), P_L reaches its global maximum when $y_j = y_k$ for all j, k for which $a_{jk} \neq 0$. If G is weakly connected, this equality propagates through the whole graph such that $y_1 = y_2 = ... = y_N$. $\square$

Proposition 6 *Consider N agents on a manifold $\mathcal{M}$ satisfying Assumptions 1 and 2. Given an undirected graph G, a local maximum of the associated cost function $P_L(y)$ necessarily corresponds to a consensus configuration and a local minimum of $P_L(y)$ necessarily corresponds to an anti-consensus configuration.*

Proof: The proof is given for local maxima; it is strictly equivalent for local minima. For $y^* = (y_1^*...y_N^*)$ to be a local maximizer of P_L, y_k^* must be, for each k, a local maximizer of $p_k(c) := P_L(y_1^*...y_{k-1}^*, c, y_{k+1}^*...y_N^*)$. Since $A = A^T$, p_k takes the linear form $p_k(c) = \xi_k + \frac{1}{N^2} c^T \left(\sum_j a_{jk} y_j^* \right)$ with ξ_k constant $\forall k$. Thanks to Assumption 2, all local maxima of $p_k(c)$ are global maxima. Therefore, y_k^* is a global maximum of $p_k(c)$ for all k, which corresponds to Definition 4 of consensus. $\square$

Proposition 6 establishes that a *sufficient* condition for (anti-)consensus configurations is to optimize P_L. However, nothing guarantees that this is also *necessary*. In general, optimizing P_L will thus provide proven (anti-)consensus configurations, but not necessarily all of them (this is because consensus maximizes P_L on $\mathcal{M}^N$ for only *moving one agent with others fixed*, and not along directions of *combined motion of several agents*). The remaining sections of this paper present algorithms that drive the swarm to (anti-)consensus. Being based on the optimization of P_L, these algorithms do not necessarily target all possible (anti-)consensus configurations. For instance, for a tree, maximization of P_L always leads to synchronization, although other consensus configurations can exist.

5.1 Examples

On $SO(n)$ and $Grass(p, n)$, P_L with matrix forms for the elements y_k becomes

$$P_L(y) = \frac{1}{2N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{jk} \operatorname{trace}(y_j^T y_k) \qquad \text{with } y_k \in \mathbb{R}^{n \times n} \ \forall k \ . \tag{11}$$

The special orthogonal group Each term $Q_j^T Q_k = Q_j^{-1} Q_k$ is itself an element of $SO(n)$. It is actually the unique element of $SO(n)$ translating Q_j to Q_k by matrix (group) multiplication on the right. Hence, on the Lie group $SO(n)$, the order parameter P_L measures the sum of the traces of the elements translating connected agents to each other. Observing that the trace is maximal for the identity matrix and considering the particular case of $SO(2)$, one can easily imagine how the trace of $Q_j^{-1} Q_k$ characterizes the distance between Q_j and Q_k. This cost function has been previously used in [8; 34] as a measure of disagreement on $SO(3)$.

The Grassmann manifold On $Grass(p, n)$, (11) can be rewritten as

$$P_L(\mathcal{Y}) = \frac{1}{2N^2} \sum_{j=1}^{N} \sum_{k=1}^{N} a_{jk} \left(\sum_{i=1}^{p} \cos^2(\phi_{jk}^i) \right)$$

with $\phi_{jk}^i = i^{\text{th}}$ principal angle between $\mathcal{Y}_j$ and $\mathcal{Y}_k$. This reformulation has previously appeared in [11; 4; 1].

6   Gradient consensus algorithms

The previous sections pave the way for ascent and descent algorithms on P and P_L. This paper considers continuous-time gradient algorithms, but any descent or ascent algorithm — in particular, discrete-time — will achieve the same task; see [2] for extensive information on this subject. In the present paper, the gradient is always defined with the canonical metric induced by the embedding of $\mathcal{M}$ in $\mathbb{R}^m$.

6.1   Fixed undirected graphs

A gradient algorithm for P_L leads to

$$\dot{y}_k(t) = 2N^2 \alpha \, \mathrm{grad}_{k,\mathcal{M}}(P_L) \,, \qquad k = 1...N \,, \tag{12}$$

where $\alpha > 0$ (resp. $\alpha < 0$) for consensus (resp. anti-consensus), $\dot{y}_k$ denotes the time-derivative of agent k's position and $\mathrm{grad}_{k,\mathcal{M}}(f)$ denotes the gradient of f with respect to y_k along $\mathcal{M}$. This gradient can be obtained from the gradient in $\mathbb{R}^m$,

$$\mathrm{grad}_{k,\mathbb{R}^m}(P_L) = \tfrac{1}{2N^2} \textstyle\sum_j (a_{jk} + a_{kj}) \, y_j \,,$$

by orthogonal projection $\mathrm{Proj}_{T\mathcal{M},k}$ onto the tangent space to $\mathcal{M}$ at y_k, yielding $\forall k$

$$\dot{y}_k(t) = \alpha \, \mathrm{Proj}_{T\mathcal{M},k} \left(\textstyle\sum_j (a_{jk} + a_{kj}) y_j \right) = \alpha \, \mathrm{Proj}_{T\mathcal{M},k} \left(\textstyle\sum_j (a_{jk} + a_{kj})(y_j - y_k) \right) . \tag{13}$$

The last equality comes from $\mathrm{Proj}_{T\mathcal{M},k}(y_k) = 0$. It shows that to implement this consensus algorithm, each agent k must know the relative position with respect to itself of all agents j such that $j \rightsquigarrow k$ or $k \rightsquigarrow j$. Since the information flow is restricted to $j \rightsquigarrow k$, (13) can only be implemented for undirected graphs, for which it becomes

$$\dot{y}_k(t) = 2\alpha \, \mathrm{Proj}_{T\mathcal{M},k} \left(\textstyle\sum_{j=1}^N a_{jk}(y_j - y_k) \right) \,, \qquad k = 1...N \,. \tag{14}$$

In the special case of a complete unit-weighted graph,

$$\dot{y}_k(t) = 2\alpha N \, \mathrm{Proj}_{T\mathcal{M},k} \left(C_e(t) - y_k \right) \,, \qquad k = 1...N \,. \tag{15}$$

Proposition 7 *A group of N agents moving according to (14) on a manifold $\mathcal{M}$ satisfying Assumptions 1 and 2, where the graph G associated to $A = [a_{jk}]$ is undirected, always converges to a set of equilibrium points. If $\alpha < 0$, all asymptotically stable equilibria are anti-consensus configurations for G. If $\alpha > 0$, all asymptotically stable equilibria are consensus configurations for G (in particular, for the equally-weighted complete graph, the only asymptotically stable configuration is synchronization).*

Proof: $\mathcal{M}$ being compact and the a_{jk} bounded, P_L is upper- and lower-bounded. P_L is always increasing (decreasing) for $\alpha > 0$ ($\alpha < 0$) along solutions of (14), since

$$\dot{P}_L \,=\, \textstyle\sum_k \dot{y}_k^T \, \mathrm{grad}_{k,\mathcal{M}}(P_L) \,=\, 2N^2\alpha \sum_k \|\mathrm{grad}_{k,\mathcal{M}}(P_L)\|^2 \,.$$

By LaSalle's invariance principle, the swarm converges towards a set where $\dot{P}_L = 0$, implying $\mathrm{grad}_{k,\mathcal{M}}(P_L) = 0 \Leftrightarrow \dot{y}_k = 0 \; \forall k$ and the swarm converges to a set of equilibria. For $\alpha > 0$ ($\alpha < 0$), since P_L always increases (decreases) along solutions, only local maxima (minima) can be asymptotically stable. Proposition 6 states that all local maxima (minima) of P_L correspond to consensus (anti-consensus).   □

Remark 1 *Computing $\mathrm{grad}_{k,\mathcal{M}}$ directly along the manifold, as in [2], can be much more efficient if the dimension of $\mathcal{M}$ is substantially lower than m (see §6.3).*

6.2 Extension to directed and time-varying graphs

Formally, algorithm (14) can be written for directed and even time-varying graphs, although the gradient property is lost for directed graphs and has no meaning in the time-varying case (since P_L then explicitly depends on time). Nevertheless, the general case of (14) with varying and directed graphs still exhibits synchronization properties.

It can be shown that synchronization is still a stable equilibrium; it is asymptotically stable if disconnected graph sequences are excluded. Its basin of attraction includes the configurations where all the agents are located in a convex set of $\mathcal{M}$. Indeed, convergence results on Euclidean spaces can be adapted to manifolds when agents are located in a convex set (see e.g. [33]). On the other hand, examples where algorithm (14) with $\alpha > 0$ runs into a limit cycle can be built for as simple cases as undirected equally-weighted (but varying) graphs on the circle (see §6.3).

Simulations on $SO(n)$ and $Grass(p, n)$ seem to indicate that for randomly generated digraph sequences[1], the swarm eventually converges to synchronization when $\alpha > 0$; this would correspond to *generic* convergence for unconstrained graphs.

Algorithm (14) can lead to a generalization of Vicsek's phase update law (see [48]) to manifolds. The Vicsek model is a discrete-time algorithm governing the headings of particles in the plane, and hence operates on the circle. It can be written as

$$y_k(t+1) \ \in \ IAM\left(\{y_j(t)|j \rightsquigarrow k \text{ in } G(t)\} \cup \{y_k(t)\}\right) , \qquad k = 1...N , \qquad (16)$$

with the definitions introduced in the present paper; interconnections among particles depend on their relative positions in the plane (so-called "proximity graphs"). Vicsek's law can be directly generalized in the form (16) to any manifold satisfying Assumption 1. Based on the previous discussions, it is clear why (16) can be viewed as a discrete-time variant of (14). When run asynchronously on a fixed undirected graph, (16) is an ascent algorithm for P_L; see [37] for a precise relationship between the continuous-time and discrete-time consensus algorithms on the circle.

6.3 Examples

Consensus on the circle is studied in [42; 37; 39; 43]; the other algorithms presented here are original.

The circle   Denoting angular positions by θ_k, the specific form of (14) for S^1 is

$$\dot{\theta}_k = \alpha' \ \sum_{j=1}^N a_{jk} \ \sin(\theta_k - \theta_j) , \qquad k = 1...N . \qquad (17)$$

For the equally-weighted complete graph, this is strictly equivalent to the Kuramoto model [26] with identical (zero) natural frequencies.

Algorithm (17) can run into a limit cycle for varying graphs. Consider a regular consensus state for an equally-weighted ring graph G_1, with consecutive agents separated by $\chi < \pi/2$ (local maximum of P_{L_1}). Define G_2 by connecting each agent to the agents located at an angle $\psi > \pi/2$ from itself with ψ properly fixed. G_2 is a collection of disconnected ring graphs and the swarm is at a local minimum of P_{L_2}. Starting the system in the neighborhood of that state and regularly switching between G_1 and G_2, the system will oscillate in its neighborhood, being driven away by G_2 and brought back by G_1 if consensus is intended and reversely if anti-consensus is intended.

---

[1]More precisely, the following distibution was examined: initially, each element a_{jk} independently takes a value in $\{0, 1\}$ according to a probability Prob(1) $= p$. The corresponding graph remains for a time t_{graph} uniformly distributed in $[t_{min}, t_{max}]$, after which a new graph is built as initially.

The special orthogonal group　The tangent space to $SO(n)$ at the identity I_n is the space of skew-symmetric $n \times n$ matrices. By group multiplication, the projection of $B \in \mathbb{R}^{n \times n}$ onto the tangent space to $SO(n)$ at Q_k is $Q_k \operatorname{Skew}(Q_k^{-1}B) = Q_k\left(\frac{Q_k^T B}{2} - \frac{B^T Q_k}{2}\right)$. This leads to the following explicit form of algorithm (14) on $SO(n)$, where the right-hand side only depends on relative positions of the agents with respect to k:

$$Q_k^{-1}\dot{Q}_k = \alpha \sum_j a_{jk} \left(Q_k^{-1}Q_j - Q_j^{-1}Q_k\right), \qquad k = 1...N . \tag{18}$$

Using Lemma 1 in the appendix, the following proves that $SO(n)$ satisfies Assumption 2. It also includes the proof of Proposition 1.

Proposition 8 *The manifold $SO(n)$ satisfies Assumption 2.*

Proof: (+ Prop.1) Consider a linear function $f(Q) = \operatorname{trace}(Q^T B)$ with $Q \in SO(n)$ and $B \in \mathbb{R}^{n \times n}$; $\operatorname{grad}_{\mathbb{R}^{n \times n}}(f) = B$ so $\operatorname{grad}_{SO(n)}(f) = \frac{Q}{2}(Q^T B - B^T Q)$. Since Q is invertible, critical points of f satisfy $(Q^T B - B^T Q) = 0$, meaning that they take the form described by Lemma 1. Using notations of Lemma 1, write $R = H\Lambda H^T$ where Λ contains the (non-negative) eigenvalues of R. This leads to

$$Q = UHJH^T \quad \Rightarrow \quad Q^T B = HJ\Lambda H^T \quad \Rightarrow \quad f(Q) = -\sum_{j=1}^{l} \Lambda_{jj} + \sum_{j=l+1}^{n} \Lambda_{jj} .$$

If $l \geq 2$, select any $m \in [2, l]$ and define $Q_\varepsilon = UHJAH^T$ where A is the identity matrix except that $A(1,1) = A(m,m) = \cos(\varepsilon)$ and $A(1,m) = -A(m,1) = \sin(\varepsilon)$ with ε arbitrarily small. It is straightforward to see that $f(Q_\varepsilon) > f(Q)$ unless $\Lambda_{11} = \Lambda_{mm} = 0$. Similarly, if $l = 1$ and $\exists\, m \geq 2$ such that $\Lambda_{mm} < \Lambda_{11}$, then $f(Q_\varepsilon) > f(Q)$ with Q_ε and A defined as previously. Therefore,

1. if $\det(B) \geq 0$, local maxima require $l = 0$ such that $Q = U$ and $f(Q)$ is the sum of the eigenvalues of R;

2. if $\det(B) \leq 0$, local maxima require U to take the form of Lemma 1 with $l = 1$ and $\Lambda_{11} \leq \Lambda_{mm} \forall m$; thus the first column of H corresponds to a smallest eigen- value of R and $f(Q)$ is the sum of $n - 1$ largest eigenvalues minus the smallest one.

　This shows that all maxima of $f(Q)$ are global maxima (since they all take the same value) and, with $B = C_e$, characterizes the *IAM*. □

The Grassmann manifold　The projection of a matrix $M \in \mathbb{S}_n^+$ onto the tangent space to $Grass(p,n)$ at Π_k is given in [29] as $\Pi_k M \Pi_{\perp k} + \Pi_{\perp k} M \Pi_k$. This leads to

$$\dot{\Pi}_k = 2\alpha \sum_j a_{jk} \left(\Pi_k \Pi_j \Pi_{\perp k} + \Pi_{\perp k} \Pi_j \Pi_k\right), \qquad k = 1...N . \tag{19}$$

In practice, the basis representation Y_k is handier than Π_k since it involves smaller matrices. Computing the gradient of $P_L(\{\Pi_k, \ k = 1...N\}) = P_L(\{Y_k Y_k^T, \ k = 1...N\})$ directly on the quotient manifold as explained in [1] leads to the algorithm

$$\dot{Y}_k = 4\alpha \sum_j a_{jk} \left(Y_j M_{j\cdot k} - Y_k M_{j\cdot k}^T M_{j\cdot k}\right), \qquad k = 1...N , \tag{20}$$

where the $p \times p$ matrices $M_{j\cdot k}$ are defined as $M_{j\cdot k} = Y_j^T Y_k$. For theoretical purposes, the projector representation is an easier choice, as for the following proofs.

Proposition 9 *The Grassmann manifold satisfies Assumption 2.*

Proof: (+ Prop.2) Consider a linear function $f(\Pi) = \text{trace}(\Pi^T B)$ where $B \in \mathbb{S}_n^+$ and Π represents $\mathcal{Y} \in Grass(p, n)$; $\text{grad}_{\mathbb{R}^{n \times n}}(f) = B$ so $\text{grad}_{Grass(p,n)}(f) = \Pi B \Pi_\perp + \Pi_\perp B \Pi$. The ranges of the first and second terms in $\text{grad}_{Grass(p,n)}(f)$ are at most $\mathcal{Y}$ and its orthogonal complement respectively, so they both equal zero at a critical point $\mathcal{Y}^*$, such that $\mathcal{Y}^*$ is an invariant subspace of B. In an appropriate basis $(e_1...e_n)$, write $\Pi^* = \text{diag}(1,...1,0,...0)$ and $B = \text{diag}(\mu_1,...\mu_p, \mu_{p+1}...\mu_n)$. If $\exists\, d \leq p$ and $l > p$ such that $\mu_d < \mu_l$, then any variation of Π^* rotating e_d towards e_l strictly increases $f(\Pi)$. Therefore, at local maxima of $f(\Pi)$, the p-dimensional space corresponding to Π must be an eigenspace of B corresponding to p largest eigenvalues. This implies that at any local maximum, $f(\Pi)$ equals the sum of p largest eigenvalues of B, so Assumption 2 is satisfied. Replacing B by C_e proves Proposition 2. $\qquad\square$

7 Consensus algorithms with estimator variables

Section 6 derives algorithms that lead to a consensus situation linked to the interconnection graph. But in many applications, the interconnection graph is just a restriction on communication possibilities, under which one actually wants to achieve a consensus for the complete graph. Moreover, allowing directed and time-varying communication graphs is desirable for robustness. This section presents algorithms achieving the same performance as those of §6 for the equally-weighted complete graph — that is, driving the swarm to synchronization or to a subset of the anti-consensus configurations for the equally-weighted complete graph which seems to contain little more than balancing — under very weak conditions on the actual communication graph. However, this reduction of information channels must be compensated by adding a consensus variable $x_k \in \mathbb{R}^m$, which interconnected agents are able to communicate, to the state space of each agent.

7.1 Synchronization algorithm

For synchronization purposes, the agents run a consensus algorithm on their estimator variables x_k in $\mathbb{R}^m$, $k = 1...N$, initialized arbitrarily but independently and such that they can take any value in an open subset of $\mathbb{R}^m$; $\forall k$, agent k's position y_k on $\mathcal{M}$ independently tracks (the projection on $\mathcal{M}$ of) x_k. This leads to

$$\dot{x}_k \;=\; \beta \sum_j a_{jk}(x_j - x_k) \qquad\qquad,\quad \beta > 0 \qquad\qquad (21)$$

$$\dot{y}_k \;=\; \gamma_S\, \text{grad}_{k,\mathcal{M}}(y_k^T x_k) = \gamma_S\, \text{Proj}_{T\mathcal{M},k}(x_k) \qquad,\quad \gamma_S > 0\,,\quad k = 1...N. \qquad (22)$$

Equation (21) is a classical consensus algorithm in $\mathbb{R}^m$, where $\dot{x}_k(t)$ points from $x_k(t)$ towards the centroid of the (appropriately weighted) $x_j(t)$ for which $j \rightsquigarrow k$ at time t. According to [33; 32; 35], if the time-varying communication graph $G(t)$ is piecewise continuous in time and uniformly connected, then all the x_k exponentially converge to a common consensus value x_∞; moreover, if $G(t)$ is balanced for all t, then $x_\infty = \frac{1}{N}\sum_k x_k(0)$ (i.e. x_∞ is the centroid of the initial x_k). This implies the following convergence property for (21),(22), where the notation IAM_g generalizes the definition (3) of the IAM when the points defining C_e are not on $\mathcal{M}$.

Proposition 10 *Consider a piecewise continuous and uniformly connected graph $G(t)$ and a manifold $\mathcal{M}$ satisfying Assumptions 1 and 2. The only stable limit configuration of the y_k under (21),(22), with the x_k initialized arbitrarily but independently and such that they can take any value in an open subset of $\mathbb{R}^m$, is synchronization at $y_\infty = \text{Proj}_{T\mathcal{M},k}(x_\infty)$; if $G(t)$ is balanced, $y_\infty = IAM_g\{x_k(0), k = 1...N\}$.*

Proof: Convergence of (21) towards $x_k = x_\infty$ $\forall k$ is proved in [32]; the property $x_\infty = \frac{1}{N}\sum_k x_k(0)$ for balanced graphs is easy to check (see [35]). As a consequence, the asymptotic form of (21),(22)

is a set of N independent systems

$$
\begin{align}
x_k &= x_\infty \tag{23}\\
\dot{y}_k &= \gamma_S \operatorname{Proj}_{T\mathcal{M},k}(x_\infty) , \qquad k = 1...N , \tag{24}
\end{align}
$$

where x_∞ is a constant. According to [30], the ω-limit sets of the original system (21),(22) correspond to the chain recurrent sets of the asymptotic system (23),(24). The first equation is trivial. According to Proposition 4 in [21] and Sard's theorem, since (24) is a gradient ascent algorithm for $f(y_k) = y_k^T x_\infty$ and $f(y_k)$ is smooth (as the restriction of a smooth function to the smooth embedded manifold $\mathcal{M}$), the chain recurrent set of (24) is equal to its critical points. Since x_∞ is a linear combination of the $x_k(0)$, variations of the $x_k(0)$ are equivalent to variations of x_∞.

Property o: Any open neighborhood O of any point $x_o \in \mathbb{R}^m$ contains a point x_a for which $f(y_k)$ has a unique (local = global, by Assumption 2) maximizer.
Proof: If $y_k^T x_o$ has multiple maximizers, select one of them, call it y_*. Then for $\sigma > 0$, $y_k^T x_o + \sigma\, y_k^T y_* \le y_*^T x_o + \sigma\, y_*^T y_* \le y_*^T x_o + \sigma\, y_*^T y_*$ with equality holding if and only if $y_k = y_*$, so y_* is the unique maximizer of $y_k^T (x_o + \sigma y_*)$. Since any open neighborhood O of x_o contains points of the form $x_a = x_o + \sigma y_*$, $\sigma > 0$, property o is proved.

Because of Property o, with respect to variations of the x_k, the situation "$f(y_k)$ has multiple maximizers" is unstable. The situation "$f(y_k)$ has a unique maximizer" is stable since it corresponds to a non-empty open set in $\mathbb{R}^m$; thus a convex neighborhood of x_∞ can be found in which the $x_k(t)$ will stay by convexity of (21) and where $f(y_k)$ has a unique maximizer. With respect to variations of the y_k, the (thus unique) maximizer of $y_k^T x_k$ is the only stable equilibrium for gradient ascent algorithm (22), such that for $x_k \to x_\infty$ the only stable situation is synchronization. $\square$

7.2 Anti-consensus algorithm

For anti-consensus, in analogy with the previous section, each y_k evolves according to a gradient algorithm to maximize its distance to $x_k(t)$. If $x_k(t)$ asymptotically converges to $C_e(t)$, this becomes equivalent to the gradient anti-consensus algorithm (15). Imposing $x_k(0) = y_k(0) \ \forall k$, the following algorithm achieves this purpose when $G(t)$ is balanced $\forall t$:

$$
\begin{align}
\dot{x}_k &= \beta \sum_{j=1}^{N} a_{jk} (x_j - x_k) + \dot{y}_k & , \quad \beta > 0 \tag{25}\\
\dot{y}_k &= \gamma_B \operatorname{grad}_{k,\mathcal{M}}(y_k^T x_k) = \gamma_B \operatorname{Proj}_{T\mathcal{M},k}(x_k) & , \quad \gamma_B < 0 , \quad k = 1...N . \tag{26}
\end{align}
$$

Note that the variables x_k and y_k are fully coupled; in a discrete-time version of this system, this essential feature of the algorithm must be retained in the form of *implicit* update equations in order to ensure convergence (see [39] for details).

Proposition 11 *Consider a piecewise continuous, uniformly connected and balanced graph $G(t)$ and a manifold $\mathcal{M}$ satisfying Assumptions 1 and 2. Then, algorithm (25),(26) with initial conditions $x_k(0) = y_k(0) \ \forall k$ converges to an equilibrium configuration of the anti-consensus algorithm for the equally-weighted complete graph, that is (15) with $\alpha < 0$.*

Proof: First show that $\frac{1}{N} \sum_k x_k(t) = \frac{1}{N} \sum_k y_k(t) = C_e(t)$. Since $x_k(0) = y_k(0) \ \forall k$, it is true for $t = 0$. Thus it remains to show that $\sum_k \dot{x}_k(t) = \sum_k \dot{y}_k(t)$. This is the case because a balanced graph ensures that the first two terms on the right side of the following expression cancel each other:

$$
\sum_k \dot{x}_k(t) = \beta \sum_j \left(\sum_k a_{jk} \right) x_j - \beta \sum_k \left(\sum_j a_{jk} \right) x_k + \sum_k \dot{y}_k(t) .
$$

Next, prove that $\forall k$, $\dot{y}_k(t)$ is a uniformly continuous function in $L_2(0, +\infty)$ such that Barbalat's Lemma implies $\dot{y}_k \to 0$. First show that $W(t) = \frac{1}{2} \sum_k x_k(t)^T x_k(t)$ is never increasing along the solutions of (25),(26). Denoting by $(x)_j$, $j = 1...m$, the vectors of length N containing the j-th

component of every x_k, $k = 1...N$ and by $L^{(i)}$ the in-Laplacian of the varying graph associated to the a_{jk}, one obtains

$$\dot{W}(t) = \sum_k x_k^T \dot{x}_k = \sum_k x_k^T \dot{y}_k - \beta \sum_j (x)_j^T L^{(i)}(x)_j .$$

The term containing $L^{(i)}$ is non-positive because the Laplacian of balanced graphs is positive semi-definite (see [49]). Replacing $\dot{y}_k$ from (26) and noting that $x_k^T \operatorname{Proj}_{T\mathcal{M},k}(x_k) = \left(\operatorname{Proj}_{T\mathcal{M},k}(x_k)\right)^T \operatorname{Proj}_{T\mathcal{M},k}(x_k)$, one obtains

$$\dot{W}(t) = \gamma_B \sum_k \|\operatorname{Proj}_{T\mathcal{M},k}(x_k)\|^2 - \beta \sum_j (x)_j^T L^{(i)}(x)_j \le 0 . \tag{27}$$

Thus $W(t) \le W(0) = \frac{N}{2} r_{\mathcal{M}}^2$ which implies that each $\dot{y}_k(t)$ is in $L_2(0, +\infty)$ since

$$\frac{1}{|\gamma_B|} \sum_k \int_0^{+\infty} \|\dot{y}_k(t)\|^2 \, dt \le - \int_0^{+\infty} \dot{W}(t) \, dt \le \frac{N}{2} r_{\mathcal{M}}^2 .$$

$W(t) \le W(0)$ also implies that x_k is uniformly bounded $\forall k$; from (26), $\dot{y}_k$ is uniformly bounded as well. Combining these two observations, with the a_{jk} bounded, (25) shows that x_k has a bounded derivative and hence is Lipschitz in t $\forall k$. Now write

$$\|\dot{y}_k(x_k(t_1), y_k(t_1)) - \dot{y}_k(x_k(t_2), y_k(t_2))\| \le$$
$$\|\dot{y}_k(x_k(t_1), y_k(t_1)) - \dot{y}_k(x_k(t_2), y_k(t_1))\| + \|\dot{y}_k(x_k(t_2), y_k(t_1)) - \dot{y}_k(x_k(t_2), y_k(t_2))\| .$$

The first term on the second line is bounded by $r_1 |t_1 - t_2|$ for some r_1 since $\dot{y}_k$ is linear in x_k and x_k is Lipschitz in t. The second term on the second line is bounded by $r_2 |t_1 - t_2|$ for some r_2 since $\dot{y}_k$ is Lipschitz in y_k (as the gradient of a smooth function along the smooth manifold $\mathcal{M}$) and $\frac{d}{dt}(y_k) = \dot{y}_k$ is uniformly bounded. Hence, $\dot{y}_k$ is Lipschitz in t and therefore uniformly continuous in t, such that Barbalat's Lemma can be applied. Therefore $\dot{y}_k \to 0$. Thus from [30], the ω-limit sets of (21),(22) correspond to the chain recurrent sets of the asymptotic system

$$\begin{aligned} \dot{x}_k &= \beta \sum_j a_{jk} (x_j - x_k) \\ 0 &= \gamma_B \operatorname{Proj}_{T\mathcal{M},k}(x_k) . \end{aligned}$$

The second line is just a static condition. The chain recurrent set of the linear consensus algorithm in the first line reduces to its equilibrium set $x_k = x_\infty$ $\forall k$. But then, from the beginning of the proof, $x_k = C_e$ $\forall k$ such that the static condition becomes $0 = \gamma_B \operatorname{Proj}_{T\mathcal{M},k}(C_e)$ $\forall k$. This is the condition for an equilibrium of anti-consensus algorithm (15) with $\gamma_B = 2\alpha N$. $\qquad\square$

In simulations, a swarm applying (25),(26) with $x_k(0) = y_k(0)$ $\forall k$ seems to generically converge to an anti-consensus configuration of the equally-weighted complete graph, that is a *stable* equilibrium configuration of (15) with $\alpha < 0$.

7.3 Examples

Applying this strategy to the circle yields the results of [39], the x_k reduce to vectors of $\mathbb{R}^2$; algorithms (22) and (26) respectively drive the y_k towards and away from the central projection of x_k onto the unit circle.

The special orthogonal and Grassmann manifolds The particular balancing algorithms will not be detailed as they are directly obtained from their synchronization counterparts. Introducing auxiliary $n \times n$-matrices X_k, (21) may be transcribed verbatim. Using previously presented expressions for $\operatorname{Proj}_{T\mathcal{M},k}(X_k)$, (22) becomes

$$\text{On } SO(n): \qquad Q_k^{-1}\dot{Q}_k = \frac{\gamma_S}{2} \left(Q_k^T X_k - X_k^T Q_k\right) \qquad , \quad k = 1...N . \tag{28}$$

$$\text{On } Grass(p, n): \qquad \dot{\Pi}_k = \gamma_S \left(\Pi_k X_k \Pi_{\perp k} + \Pi_{\perp k} X_k \Pi_k\right) , \quad k = 1...N . \tag{29}$$

Note that for $Grass(p, n)$, the projector representation must be used in (21) and (25), such that using $n \times n$ matrices X_k becomes unavoidable.

7.4   Remark about the communication of estimator variables

To implement the algorithms of this section, interconnected agents must communicate the values of their estimator variable x_k. It is important to note that the variables x_k may not just be a set of abstract scalars for each agent k: since x_k interacts with the geometric y_k, it must be a geometric quantity too. However, the x_k evolve in $\mathbb{R}^m$ while the original system lives on $\mathcal{M}$; the relative position of agents on $\mathcal{M}$ is a meaningful measurement, but nothing ensures *a priori* that a similar thing can be done in $\mathbb{R}^m$. A solution could be to use a common (thus external) reference frame in $\mathbb{R}^m$ and transmit the coordinates of the x_k in this frame. That solution would unfortunately imply that the swarm loses its full autonomy; however, the external frame is just used for "translation" purposes and does not interfer with the dynamics of the system.

When $\mathcal{M}$ is (a subgroup of) $SO(n)$, the algorithms can be reformulated such that they work completely autonomously if interconnected agents measure their relative positions $Q_k^T Q_j$. Indeed, define $Z_k = Q_k^T X_k$. Then (21),(22) for instance becomes

$$\dot{Z}_k \;\; = \;\; (Q_k^T \dot{Q}_k)^T Z_k \; + \; \beta \sum_j a_{jk} \left((Q_k^T Q_j) Z_j - Z_k \right) \tag{30}$$

$$Q_k^T \dot{Q}_k \;\; = \;\; \tfrac{\gamma_S}{2} \left(Z_k - Z_k^T \right) \qquad\qquad , \; k = 1...N \; . \tag{31}$$

In this formulation, each agent k can represent Z_k as an array of scalars, whose columns express the column-vectors of X_k as coordinates in a local frame attached to k (i.e. in a frame rotated by Q_k with respect to a hypothetical reference frame). Pre-multiplying Z_j by $Q_k^T Q_j$ expresses X_j in the local frame of k, and $Q_k^T \dot{Q}_k$ expresses the velocity of Q_k (with respect to a hypothetical fixed reference) in the local frame of k as well. Thus (30),(31) actually corresponds to (21),(22) written in the local frame of k. Each agent k gets from its neighbors $j \rightsquigarrow k$ their relative positions $Q_k^T Q_j$ and the $n \times n$ arrays of numbers Z_j; from this it computes the update $\dot{Z}_k$ to its own array of numbers Z_k and the move it has to make with respect to its current position, $Q_k^T \dot{Q}_k$. The same can be done for the anti-consensus algorithm.

8   Conclusion

The present paper makes three main contributions.

First, it defines the induced arithmetic mean of N points on an embedded connected compact homogeneous manifold $\mathcal{M}$; though it differs from the traditional Karcher mean, it has a clear geometric meaning with the advantage of being easily computable — see analytical solutions for $SO(n)$ and $Grass(p, n)$.

Secondly, a definition of consensus directly linked to the induced arithmetic mean is presented for these manifolds. In particular, the notion of balancing introduced in [42] for the circle is extended to connected compact homogeneous manifolds. Consensus for the equally-weighted complete graph is equivalent to synchronization. Likewise, it appears in simulations that anti-consensus for the equally-weighted complete graph leads to balancing (if N is large enough), even though this could not be proved.

Thirdly, consensus is formulated as an optimization problem and distributed consensus algorithms are designed for N agents moving on a connected compact homogeneous manifold. In a first step, gradient algorithms are derived for fixed undirected interconnection graphs; (anti-)consensus configurations are their only stable equilibria. Similar algorithms are considered when the graph is allowed to be directed and/or to vary, but their convergence properties are mostly open. In a second step, the algorithms are modified by incorporating an estimator variable for each agent. In this setting, convergence to the (anti-)consensus states of the equally-weighted complete graph can be established theoretically for time-varying and directed interconnection graphs. The meaningful way of communicating estimators between agents remains an open issue when $\mathcal{M}$ is not a subgroup of $SO(n)$.

Running examples $SO(n)$ and $Grass(p, n)$ illustrate the validity of the discussion and provide geometric insight. The models and results obtained by applying this framework to the circle are

strictly equivalent to existing models and results (most significantly in [42],[43],[39]). This draws a link from the present discussion to the vast literature about synchronization and balancing on the circle.

9 Appendix

Lemma 1 *If $g(Q) = Q^T B - B^T Q$ with $Q \in SO(n)$ and $B \in \mathbb{R}^{n \times n}$, then $g(Q) = 0$ iff $Q = UHJH^T$, where $B = UR$ is a polar decomposition of B, the columns of H contain (orthonormalized) eigenvectors of R and*

$$J = \begin{pmatrix} -I_l & 0 \\ 0 & I_{n-l} \end{pmatrix}, \qquad \begin{array}{ll} l \ even & if \ \det(U) > 0 \\ l \ odd & if \ \det(U) < 0 \end{array}$$

Proof: All matrices Q of the given form obviously satisfy that $Q^T B$ is symmetric. The following constructive proof shows that this is the only possible form.

Since $U^T B = R$ is symmetric with $U \in O(n)$, the problem is to find all matrices $T = U^T Q \in O(n)$ such that $S = T^T R$ is symmetric and $\det(T) = \det(U)$. Work in a basis of eigenvectors H^* diagonalizing R with its eigenvalues placed in decreasing order $\lambda_1 \geq \lambda_2 ... \geq \lambda_n \geq 0$. The following shows that T is diagonal in that basis. Then orthogonality of T imposes values 1 or -1 on the diagonal, the number l of -1 being compatible with $\det(T) = \det(U)$; the final form follows by returning to the original basis and reordering the eigenvectors such that those corresponding to -1 are in the first columns.

The jth column of S is simply the jth column of T multiplied by λ_j. Therefore:

1. If $\lambda_i = \lambda_j$, then H^* may be chosen such that the corresponding submatrix $T(i:j, i:j) =$ intersection of rows i to j and columns i to j of T is diagonal.

2. If $\lambda_{p+1} = 0$ and $\lambda_p \neq 0$, then S symmetric implies $T(n-p:n, 1:p) = 0$. As $T(n-p:n, n-p:n)$ is diagonal from 1., only diagonal elements are non-zero in the last $n-p$ rows of T. Rows and columns of T being normalized, $T(1:p, n-p:n) = 0$.

3. Consider $i_- \leq p$ and i_+ the smallest index such that $\lambda_{i_+} < \lambda_{i_-}$. Note that

$$\sum_j T_{i_- j}^2 = \sum_j T_{j i_-}^2 = 1 \text{ (orthogonality) and } \sum_j S_{i_- j}^2 = \sum_j S_{j i_-}^2 \text{ (symmetry).} \qquad (32)$$

Start with $i_- = 1$ and assume $\lambda_{i_+} > 0$. (32) can only be satisfied if $T_{jk} = T_{kj} = 0 \ \forall j \geq i_+$ and $\forall k \in [i, i_+)$; 1. further implies $T_{jk} = T_{kj} = 0 \ \forall j \neq k$ and $\forall k \in [i_-, i_+)$. This argument is repeated by defining the new i_- as being the previous i_+ until $\lambda_{i_+} = 0$ (case 2.) or $\lambda_{i_-} = \lambda_n > 0$. This leaves T diagonal. $\qquad\square$

References

[1] P.A. Absil, R. Mahony, and R. Sepulchre. Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. *Acta Applicandae Mathematicae*, 80(2):199–220, 2004.

[2] P.A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds.* Princeton University Press, 2007.

[3] A. Barg. Extremal problems of coding theory. In H. Niederreiter, editor, *Coding Theory and Cryptography*, Notes of lectures at the Institute for Math. Sciences of the University of Singapore, pages 1–48. World Scientific, 2002.

[4] A. Barg and D.Y. Nogin. Bounds on packings of spheres in the Grassmann manifold. *IEEE Transactions on Information Theory*, 48(9):2450–2454, 2002.

[5] A.K. Bondhus, K.Y. Pettersen, and J.T. Gravdahl. Leader/follower synchronization of satellite attitude without angular velocity measurements. *Proc. 44th IEEE Conf. on Decision and Control*, pages 7270–7277, 2005.

[6] R.W. Brockett. Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems. *Linear Algebra and its Applications*, 146:79–91, 1991.

[7] M. Brookes. Matrix reference manual. *Imperial College London*, 1998-2005.

[8] F. Bullo and R. Murray (advisor). Nonlinear control of mechanical systems: a Riemannian geometry approach. *PhD Thesis, CalTech*, 1998.

[9] S.R. Buss and J.P. Fillmore. Spherical averages and applications to spherical splines and interpolation. *ACM Transactions on Graphics*, 20(2):95–126, 2001.

[10] F.R.K. Chung. *Spectral Graph Theory*. Number 92 in Regional Conference Series in Mathematics. AMS, 1997.

[11] J.H. Conway, R.H. Hardin, and N.J.A. Sloane. Packing lines, planes, etc.: Packings in Grassmannian spaces. *Exper.Math.*, 5(2):139–159, 1996.

[12] J. Cortes, S. Martinez, and F. Bullo. Coordinated deployment of mobile sensing networks with limited-range interactions. *Proc. 43rd IEEE Conf. on Decision and Control*, pages 1944–1949, 2004.

[13] T.D. Downs. Orientation statistics. *Biometrika*, 59:665–676, 1972.

[14] A. Edelman, T.A. Arias, and S.T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, 1999.

[15] G.A. Galperin. A concept of the mass center of a system of material points in the constant curvature spaces. *Comm. Math. Phys.*, 154:63–84, 1993.

[16] D. Groisser. Newton's method, zeroes of vector fields, and the Riemannian center of mass. *Adv. Appl. Math.*, 33:95–135, 2004.

[17] P. Gruber and F.J. Theis. Grassmann clustering. *Proc. 14th European Signal Processing Conference*, 2006.

[18] U. Helmke and J.B. Moore. *Optimization and dynamical systems*. Springer, 1994.

[19] J.J. Hopfield. Neural networks and physical systems with emergent collective computational capabilities. *Proc. Nat. Academy of Sciences*, 79:2554–2558, 1982.

[20] K. Hueper and J. Manton. The Karcher mean of points on $SO(n)$. *Talk at Cesame (UCL, Belgium)*, 2004.

[21] M. Hurley. Chain recurrence, semiflows, and gradients. *Journal of Dynamics and Differential Equations*, 7(3):437–456, 1995.

[22] A. Jadbabaie, J. Lin, and A.S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Transactions on Automatic Control*, 48(6):988–1001, 2003.

[23] E. Justh and P. Krishnaprasad. A simple control law for UAV formation flying. Technical report, TR 2002-38, ISR, University of Maryland, 2002.

[24] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure and Applied Math.*, 30:509–541, 1977.

[25] T.R. Krogstad and J.T. Gravdahl. Coordinated attitude control of satellites in formation. In *Group Coordination and Cooperative Control*, volume 336 of *Lecture Notes in Control and Information Sciences*, chapter 9, pages 153–170. Springer, 2006.

[26] Y. Kuramoto. In *Internat. Symp. on Math. Problems in Theoretical Physics*, volume 39 of *Lecture Notes in Physics*, page 420. Springer, 1975.

[27] J.R. Lawton and R.W. Beard. Synchronized multiple spacecraft rotations. *Automatica*, 38:1359–1364, 2002.

[28] N.E. Leonard, D. Paley, F. Lekien, R. Sepulchre, D. Frantantoni, and R. Davis. Collective motion, sensor networks and ocean sampling. *Proceedings of the IEEE*, 95(1):48–74, January 2007.

[29] A. Machado and I. Salavessa. Grassmannian manifolds as subsets of Euclidean spaces. *Res. Notes in Math.*, 131:85–102, 1985.

[30] K. Mischaikow, H. Smith, and H.R. Thieme. Asymptotically autonomous semi-flows, chain recurrence and Lyapunov functions. *Transactions of the AMS*, 347(5):1669–1685, 1995.

[31] M. Moakher. Means and averaging in the group of rotations. *SIAM J. Matrix Anal. Appl.*, 24(1):1–16, 2002.

[32] L. Moreau. Stability of continuous-time distributed consensus algorithms. *Proc. 43rd IEEE Conf. on Decision and Control*, pages 3998–4003, 2004.

[33] L. Moreau. Stability of multi-agent systems with time-dependent communication links. *IEEE Transactions on Automatic Control*, 50(2):169–182, 2005.

[34] S. Nair and N.E. Leonard. Stabilization of a coordinated network of rotating rigid bodies. *Proc. 43rd IEEE Conf. on Decision and Control*, pages 4690–4695, 2004.

[35] R. Olfati-Saber and R.M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on Automatic Control*, 49(9):1520–1533, 2004.

[36] X. Pennec. Probabilities and statistics on Riemannian manifolds: a geometric approach. *INRIA research report*, 5093, 2004.

[37] A. Sarlette, R. Sepulchre, and N.E. Leonard. Discrete-time synchronization on the N-torus. *Proc. 17th Intern. Symp. on Mathematical Theory of Networks and Systems*, pages 2408–2414, 2006.

[38] A. Sarlette, R. Sepulchre, and N.E. Leonard. Cooperative attitude synchronization in satellite swarms: a consensus approach. *Proc. 17th IFAC Symp. on Automatic Control in Aerospace*, 2007.

[39] L. Scardovi, A. Sarlette, and R. Sepulchre. Synchronization and balancing on the N-torus. *Systems and Control Letters*, 56(5):335–341, 2007.

[40] L. Scardovi and R. Sepulchre. Collective optimization over average quantities. *Proc. 45th IEEE Conf. on Decision and Control*, 2006.

[41] R. Sepulchre, D. Paley, and N.E. Leonard. Group coordination and cooperative control of steered particles in the plane. In *Group Coordination and Cooperative Control*, volume 336 of *Lecture Notes in Control and Information Sciences*, chapter 13, pages 217–232. Springer, 2006.

[42] R. Sepulchre, D. Paley, and N.E. Leonard. Stabilization of planar collective motion with all-to-all communication. *IEEE Transactions on Automatic Control*, 52(5):811–824, 2007.

[43] R. Sepulchre, D. Paley, and N.E. Leonard. Stabilization of planar collective motion with limited communication. *IEEE Transactions on Automatic Control*, to appear.

[44] S.H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled nonlinear oscillators. *Physica D*, 143:1–20, 2000.

[45] S.H. Strogatz. *Sync: The emerging science of spontaneous order*. Hyperion, 2003.

[46] J.N. Tsitsiklis and D.P. Bertsekas. Distributed asynchronous optimal routing in data networks. *IEEE Transactions on Automatic Control*, 31(4):325–332, 1986.

[47] J.N. Tsitsiklis, D.P. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Transactions on Automatic Control*, 31(9):803–812, 1986.

[48] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75(6):1226–1229, 1995.

[49] J. C. Willems. Lyapunov functions for diagonally dominant systems. *Automatica*, 12:519–523, 1976.